

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.



FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN

101900-57-L
Page 2

Developing Processing Techniques for Skylab Data
Monthly Progress Report, May 1975

The following report serves as the twenty-seventh monthly progress report for EREP Investigation 456 M which is entitled "Developing Processing Techniques for Skylab Data". The financial report for this contract (NAS9-13280) is being submitted under separate cover.

The purpose of this investigation is to test information extraction techniques for SKYLAB S-192 data and compare with results obtained in applying these techniques to LANDSAT and aircraft scanner data.

Progress on this contract was impeded for much of May by ERIM's relocation to different facilities. [ERIM has purchased a large laboratory-office building in Ann Arbor, and the organization is relocating to the new site.] During the period from May 8 to May 26 the ERIM computer facility was being moved and reestablished in the new building and consequently was closed during this period. Work after May 8, therefore, concentrated on those analysis tasks that did not require the use of the computer.

During the month work continued on the analyses of the effects of misregistration in SKYLAB data. Work was also begun on the analysis of the signature set derived from the S-192 multispectral data for suitability for use with the subresolution element classifier.

SIGNIFICANT RESULTS:

DETERMINATION OF SPATIAL MISREGISTRATION

We had previously analyzed spatial misregistration effects for conic (not scan-line-straightened) format data and showed the existence and extent of misregistration in it. This month we analyzed the scan-line-straightened data for spatial misregistration. It is a complex topic and the confusion involved in understanding it rises exponentially the deeper one tries to study it.

Let's begin the explanation in a simple fashion. First, we will define misregistration as the distance between the centers of the effective resolved areas in two channels (SDOs) of one pixel.

N75-27527

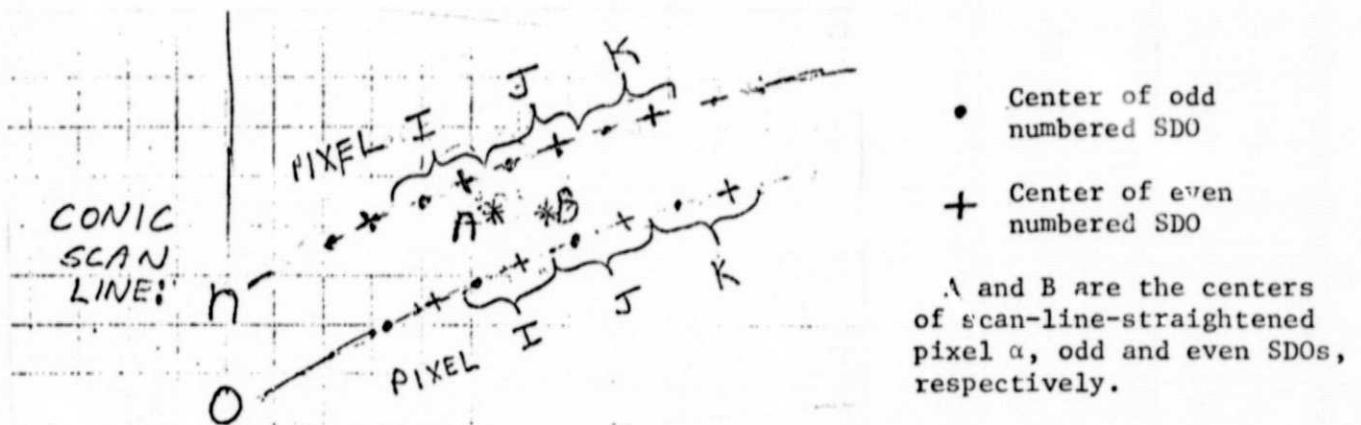
Unclas
00339

G3/43

(E75-10339) DEVELOPING PROCESSING
TECHNIQUES FOR SKYLAB DATA Monthly Progress
Report, May 1975 (Environmental Research
Inst. of Michigan) 11 p HC \$3.25 CSCL 05B

By the S-192 system design, all even numbered SDOs are perfectly registered one with the other; the same is true for all odd number SDOs. Further, there is a one-half pixel misregistration between the odd numbered SDOs and the even numbered SDOs due to the sampling technique used. Further misregistration is introduced by the scanner electronics, by different response times for different detectors, and/or by improperly skewed record heads on the spacecraft tape recorder. These combine to produce the misregistration observed in the conic data. Previous reports showed that these effects are constant over a single scan line and over a large set of scan lines and subsequently could be measured. Also, remedial efforts can be made by shifting SDOs relative to pixel numbers where the misregistration is close to a multiple of 1 pixel.

When the scan-line-straightening algorithm rearranges the collected pixels into scan-line-straightened format, additional spatial misregistration is introduced. The following example gives a graphic account of the randomness of the resulting misregistration and the possible extent of it. In what follows we present two pixels each from two conic scan lines and show the manner in which they are assigned to a straightened scan line.



To begin the analyses, let us break the 22 SDOs into four subsets and examine each independently. In what follows we will assume that all SDOs in a subset will be assigned in the same way; this is accurate since the assignment algorithm is the same for all SDOs and the starting point on a scan line is the same for all SDOs.

All ODD numbered, LOW sample rate SDOs from pixel j in line n will be assigned to A of scan-line-straightened pixel (SLSP) α . Similarly all EVEN numbered LOW sample rate SDOs from pixel i , scan line 0 will be assigned to B. (B being the even numbered SDOs of pixel α .)

All EVEN numbered, HIGH sample rate SDOs from pixel i , scan line n , will be assigned to A and renamed to be the ODD numbered SDOs of SLSP α . Similarly all ODD numbered HIGH sample rate SDOs from pixel j scan line 0 will be assigned to B and become the EVEN numbered SDOs for SLSP α .

In each case cited above, the low sample rate and the high sample rate groups, the misregistration between the even SDOs and the odd SDOs will be that as found in the conic data -- for the along scan line direction. In the along track direction there will be one full pixel misregistration.

The misregistration between a set of high sample rate SDOs and a set of low sample rate SDOs is indeterminate since it depends on whether or not the even-odd designation for the high sample rate SDOs in the straightened format has been switched from what it was in the conic format.

Last month we discussed the need to define an inset factor as an aid in identifying field center pixels for fields whose boundaries are known. Such an inset accounts for various errors and effects in the data to insure that for the field center pixels identified, all SDOs of those pixels are resolving pure field center areas. We introduced an equation with five components to calculate the inset. The contribution to the total inset due to misregistration effects was referred to as R_x , for the component along the straightened scan line, and R_y for the along track component. From examples such as that cited above, it is possible to calculate R_x and R_y as follows:

$$R_x = 1 + M \sin \theta \quad (\text{pixels})$$

$$R_y = 1 + M \cos \theta \quad (\text{pixels})$$

where

M is the maximum misregistration in the conic data

and

θ is the angle between a line drawn tangent to the conic scan at the scan point of the field of interest and a line drawn perpendicular to the straightened scan line.

If the computer program used to implement the insets cannot calculate a new inset for each point but uses a constant inset, then

$$R_x = R_y = 1 + M$$

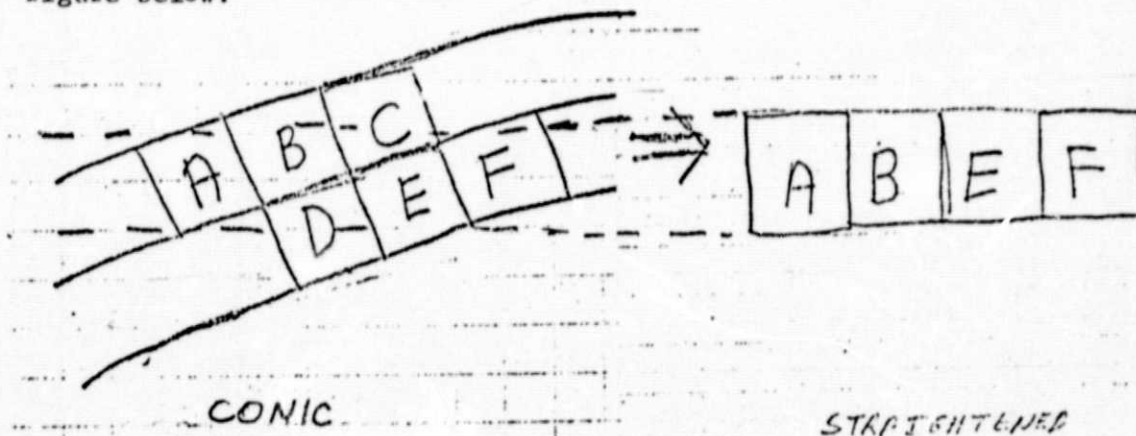
and for this data set, $M = .82$ so

$$R_x = R_y = 1.82 \text{ pixels}$$

would have to be used. Returning to last month's discussion we find that the total inset needed for identifying field center pixels in scan line straightened data would be:

$$I = I_x = I_y = 3.4 \text{ pixels}$$

Another observation regarding misregistration in scan line straightened data is that it is not possible to correct the data, at least not using a simple algorithm as was used in the conic data. Further, it is not possible even to correct within any one of the four subsets previously cited, so that misregistration due to scanner electronics could be reduced even within one of subsets of SDOs. That this is the case may be easily shown by using the figure below.



In the figure pixels A, B, E, and F will be assigned sequentially to a straightened scan line. Assume that one SDO, SDO k, is one pixel out of registration with the other SDOs. Thus SDO k of pixel B images the area of pixel A, and SDO k of pixel E images the area of pixel D. Any attempt

to simply shift SDO k one pixel relative to all the other SDOs will result in SDO k of pixel E being the area of pixel B, and not pixel D as would be correct. It is possible that such a technique would reduce the misregistration in some pixels, but it would increase the misregistration for other pixels and, more importantly, it would not be possible to know exactly which pixels were correct and which ones weren't.

Finally, it is obvious that the increased misregistration caused by the scan-line-straightening algorithm results in fewer pure field center pixels and in many more pseudo mixture pixels, i.e., pixels which have some SDOs imaging field center areas and other SDOs imaging field boundaries or even completely different fields.

The effects of misregistration and the scan-line-straightening algorithm on S-192 data may be stated succinctly:

1. There is greatly increased misregistration in scan-line-straightened data over conic data.
2. Scanner caused misregistration between any pairs of channels may not be corrected for in scan-line-straightened data.
3. Scan-line-straightened data will have fewer pure field center pixels than will conic data.

EFFECTS OF CHANNEL-TO-CHANNEL SPATIAL MISREGISTRATION ON CLASSIFICATION ACCURACY AND ON PROPORTION ESTIMATION

Previous reports describe two simulation techniques developed to investigate the effects of channel-to-channel misregistration on recognition accuracy. An experiment was also proposed to investigate the effects of misregistration on field center and border pixel classification. During this report period a program SIMSIG was developed implementing the signature simulation model; data processing stages of the proposed experiment were carried out, and an analysis was made of the effects of spatial misregistration on field center classification accuracy. A discussion of this analysis follows.

Insight was gained into what effects spatial misregistration may have on field-center recognition performance first through an analytical analysis of the problem. This analysis examined two normal distributions with common covariance for any number of channels of data. The conclusions of the analysis were intriguing. Where 'common sense' might dictate the hypothesis

that misregistration would hurt field-center recognition performance, the model studied indicated that quite the opposite could be true. Under certain circumstances misregistration could actually improve results in the classification of field center pixels!

Since misregistration and correlation are highly related, the analysis examined error rate of classification as a function of correlation (ρ). It was determined that a unique maximum error rate is reached somewhere between $-1 \leq \rho \leq 1$. Figure 1 plots error rate ϕ as a function of correlation ρ in a conceivable manner as determined by the analysis. Misregistering data will cause correlation to tend to zero. Therefore, should the given correlation ϕ between the two stated distributions lie in the range $0 \leq \phi \leq \rho_{\text{crit}} \leq 1$ for perfectly registered data, then by misregistering the data the expected error rate would actually decrease in value.

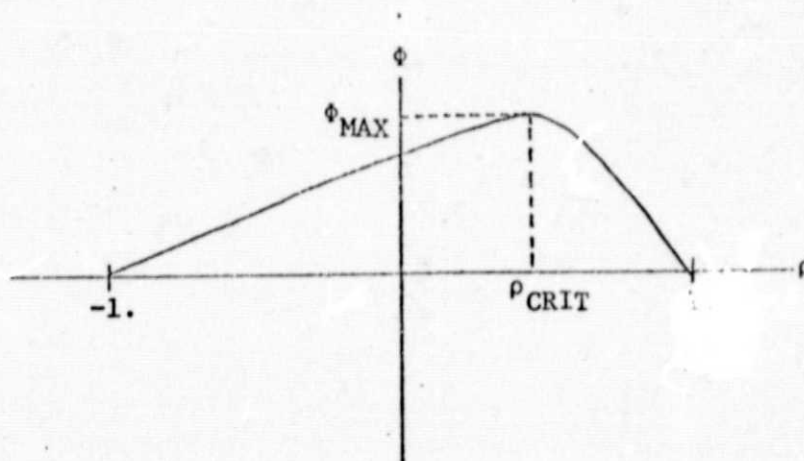


FIGURE 1. ERROR RATE OF RECOGNITION ϕ AS A FUNCTION OF CORRELATION ρ IN FIELD CENTERS

In order to test this hypothesis in a real data processing situation, five Skylab field center signatures representing the distributions of tree, corn, grass, bare soil and brush classes were selected for use in the implementation of the experiment proposed last month. Using the algorithm discussed

in the last monthly report, signatures representing field center distributions misregistered by factors of $1/3$, $1/2$, $2/3$ and 1 whole pixel in the SDOs 2, 12 and 17 were simulated using the program SIMSIG. These three SDOs were chosen because they were found to be the three best channels for purposes of discrimination for the given signature set. An expected performance matrix was calculated for each of the four sets of simulated signatures along with the original signature set using the program PEC mentioned in the April monthly. Table 1 displays several of the calculations made.

TABLE 1. EXPECTED PERFORMANCE OF SKYLAB SIGNATURES FOR VARYING DEGREES OF MISREGISTRATION

<u>Degree of Misregistration</u>	<u>Expected Recognition Accuracy</u>					
	Tree	Grass	Bare	Brush	Corn	Total
0 pixels	96.5	81.3	97.9	77.2	77.9	86.16
$1/3$	96.3	80.3	98.2	76.8	78.2	85.96
$1/2$	96.1	81.1	98.1	76.0	78.4	85.94
$2/3$	96.2	81.8	98.1	76.0	79.0	86.22
1	96.7	82.8	98.7	76.4	78.4	86.60

The results displayed in this table seem to support the unexpected hypothesis that misregistration need not be harmful to the recognition performance of field center pixels. Note that in the total expected classification for the given signature set, for misregistration in three channels of $1/2$ a pixel the performance diminishes slightly but as more misregistration is introduced, the performance improves somewhat.

These results, however, should not suggest using misregistered data or actually misregistering data to improve recognition. This experiment indicates that while spatial misregistration may aid in the classification of pure field center pixels, two serious problems arise: (1) misregistering data has the effect of greatly decreasing the number of pure field center pixels; and (2) the classification of mixture pixels will be adversely affected especially causing a greater number of false alarms. The second

effect is still in the hypothesis stage. The analysis of this coming month will test this, centering on the effects of spatial misregistration of border pixels. To date, data has been processed in the manner proposed in the last monthly. One parameter that had been fixed was the number of channels to be misregistered. The three best channels have been misregistered in simulation and further processing will be carried out in which only the one best channel will be misregistered. This additional information will enable us to examine to a degree the effects of channel-to-channel misregistration as a function of the number of channels out of registration.

GEOMETRIC ANALYSIS OF THE SIGNATURE SET

In order to use the subresolution element classifier, or mixtures classifier as we tend to call it, it is necessary that the signatures being used are sufficiently distant in n-space one from the others, otherwise the algorithm breaks down. It is also necessary to minimize the number of signatures used for classifying in order to minimize costs. The first order of business then was to identify a small subset of signatures and then to analyze them for separation.

An analysis of the 15 signatures originally used for classifying the data showed the following breakdown:

CORN	4 Signatures
TREES	2 Signatures
BRUSH	1 Signature
GRASSES, WEEDS, ETC.	5 Signatures
BARE SOIL	1 Signature
SOYBEANS	1 Signature
ALFALFA	1 Signature

Since soybeans and alfalfa are very minor ground covers in the test site, we excluded them from this study. An analysis of the tree and brush signatures showed the two tree signatures to be very disparate, but that the brush and one of the tree signatures were very similar spectrally -- overlapping some 75%. The brush signature, representing primarily areas of scrub forest, was therefore combined with the one tree signature. As for the corn signatures, the two signatures with most of the corn points were found to be very different; since corn is a major cover, both these signatures were selected for use. The bare soil signature was included.

For the grasses, rather than combining signatures, resulting in a signature with a very large spread, we endeavored to choose just one signature. From an examination of 2-dimensional scatter plots of all the signatures we identified one grass signature which was always more toward the vegetative vertex of the so called cluster pattern. We selected this grass signature reasoning that pixels from pasture or weed fields would most likely be called a mixture of the grass and bare soil.

The signature set thus identified was used as an input to program GEOM which calculates measures of separateness for a signature simplex. The metric calculated is roughly the distance between the mean of one signature and the hyperplane through the other signatures expressed in standard deviations. If the metric calculated for a given signature is small, then the simplex is too degenerate and the mixtures algorithm will not work well.

The table below, Table 2, lists the GEOM output for each pair of signatures from the set. We consider this output because in the current implementation of the mixture classifier we will find the most likely mixture of two classes for each pixel. In general, a GEOM number greater than 2.5 is considered to indicate a diverse enough signature set to allow the running of the mixtures classifier. For the most part the numbers in the table are well above this level. This is taken to be an encouraging result.

During the coming month we will continue our analysis of the signatures and their implementation on the mixtures classifier.

TABLE 2. GEOM OUTPUTS FOR 6 SIGNATURE SET
(Read RESULT for Row Signature to Hyperplane of Column Entry)

	CORN 1	CORN 4	GRASS 2	BARE SOIL	SCRUB FOREST	TREE 3
CORN 1	-	2.8	4.6	6.0	2.9	3.8
CORN 4	3.6	-	6.5	11.6	1.9	4.2
GRASS 2	8.2	7.5	-	16.6	7.7	15.6
BARE SOIL	8.2	9.8	9.0	-	8.9	8.0
SCRUB FOREST	2.0	1.6	3.9	6.7	-	3.6
TREES	7.4	10.3	14.9	10.0	9.6	-

101900-57-L

Page 11

Submitted by:

Richard F. Nalepka

Richard F. Nalepka
Principal Investigator

Approved by:

Jon D. Erickson

Jon D. Erickson
Head, Information Systems
and Analysis Department

Approved by:

Paul R. Legault for

Richard R. Legault
Director, Infrared and
Optics Division

dd

